

Building Safer Autonomous Agents by Leveraging Risky Driving Behavior Knowledge

Ashish Rana*

Department Of Computer Science & Engineering
Thapar Institute of Engineering And Technology
Patiala, India
arana_be15@thapar.edu

Avleen Malhi

Department of Computing & Informatics
Bournemouth University
Bournemouth, UK
amalhi@bournemouth.ac.uk

Abstract—The highway-env reinforcement learning tasks provides a good abstract testbed for designing driving agents for specific driving scenarios like lane changing, parking or intersections etc. But, generally these driving simulation environments often restrict themselves to safer and precise trajectories. However, we clearly know that real driving tasks often involve very high risk collision prone unexpected situations. Hence, the autonomous model-free driving agents prepared in these environments are blind to certain low probability traffic collision corner cases. In our study we systematically focus on generating adversarial driving collision prone scenarios with dangerous driving behavior and heavy traffic in order to create robust autonomous agents. In our experimentation we train model free learning agents with additional collision prone scenario simulations and compare their efficacy with regular simulation based agents. Ultimately, we create a causal experimentation setup which successfully accounts for the performance improvements across different driving scenarios by utilizing learning from risky driving situations.

Index Terms—Autonomous Agents, Driving Simulations, Trajectory Prediction, Causal Analysis

I. INTRODUCTION

1

Autonomous driving agents have transformed the automobile industry into a new digital and intelligent era. Further, it will be responsible for paving the way of new safer and efficacious automotive innovations as well. Current industry progression trends show that individual autonomous agents will dominate before a federated fleet of connected autonomous agents can operate [1]. Connected autonomous agent society is still far from the truth because of safety infrastructure, security and public policy reasons [2]–[4]. Therefore, our immediate focus should be on making the driving vehicles safe and efficient. Creating safer and robust agents have already been explored quite well for the past few years as it is very important to know the agent behavior in critical scenarios [5]–[9]. In regular driving situations risk prone scenarios don't happen frequently which makes the learning process harder because of the sparsity in number of risky interactions [10]. Also, it is unethical to create risky driving scenarios for experimental studies. Therefore, generating and studying these scenarios is a very hard task.

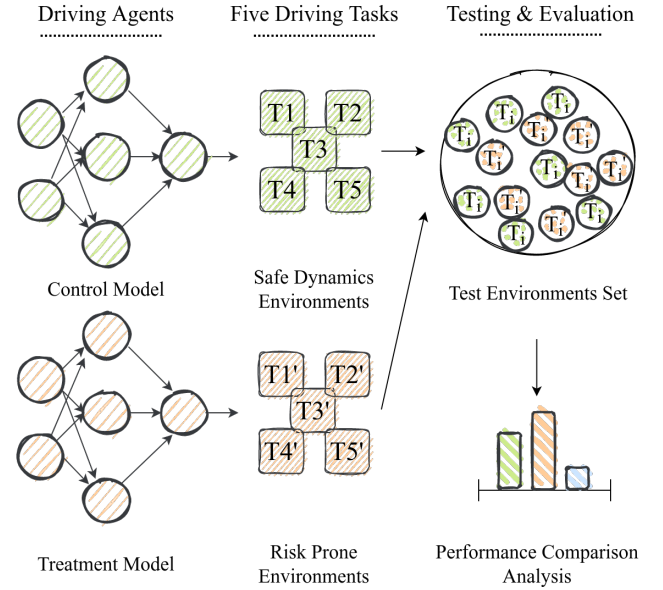


Fig. 1. Control & Treatment agents evaluated randomly for conservative and collision prone test driving environments.

Noisy labelling adversarial approaches in perception systems do demonstrate promising future research paths [11]–[13]. But the important focus remains on the constant interactions with other vehicles on roads which themselves operate independently and unexpectedly. Simulation environments with good system dynamic approximations do serve as great testbeds to simulate different critical driving tasks. Here, in this paper we formulate the driving system dynamics as Markov Decision Process (MDP) for our agent prototyping and testing tasks [14]. In our study we model our value function approximation autonomous agents with deep reinforcement learning architectures for the defined MDP driving tasks.

In our experiment setup we create two unique variants of the deep reinforcement learning model architecture for the highway-env simulation tasks like lane changing, performing u-turn etc. One of these two model variants is trained on increased dangerous driving scenario interventions for the driving related tasks in the highway-env package. Whereas the second model variant represents the control variable in our

experiment which is used to define the baseline vanilla agent trained on regular conservative simulations. In this study we perform systematic doping of these environments with random dangerous driving scenarios to create more collision prone environments. This is done in two ways, first we increase randomization behavior and second we increase traffic at key locations making the driving maneuvers really hard to perform. These two factors combined together makes the driving environment more collision prone for our ego-vehicle agent. This is done to create more robust agents in a given environment for that particular scenario. The intent is to make these agents better at post impact trajectory predictions of other colliding vehicles. Figure 1 elaborates the causal analysis aspect of the experimentation setup in randomized control trial research design.

As we hold complete control of the data generation process in our reinforcement learning setup we also attempt to understand our experimentation setup from a causal viewpoint. This in turn provides us with a unique vantage point of designing the experiment study in the form of a randomized control trial (RCT). We train our autonomous agents with a strong assumption of absence in unobservable confounding variables as we have strictly defined the vehicle behavior governing models and state dynamics. The customization features in the highway-env simulation package gives us control to keep every condition the same except for our treatment. Essentially, meaning that there is a covariate balance and comparability in our experiment design [16]–[18]. This special relationship allows us to imply that association found while calculating potential outcomes is causation. Figure 2 demonstrates the treatment (T) is subjecting the agent learning process to risk prone collision scenarios in a given driving environment task. The sample test experiment population includes evaluating the two model variants against the regular and perturbed environments with additional collision prone corner cases. Finally, by utilizing the expectation equations derived from the causal structure graph provided in Figure 2 we estimate the causal impact of novel learning changes for enhanced efficacy and safety.

Our main contribution from this paper includes providing an experimentation methodology for reinforcement learning modeling techniques that helps in calculating the causal impacts of our treatments. Along with that this methodology assists in creating more robust agents with better on-road safety. Also, we benchmark performance in simulation driving tasks for both treatment and control agents. Further in this paper we first discuss the previous work corresponding to autonomous simulation driving agents and utilization of risk prone behavior for creating safer autonomous vehicles. After that we formally define the reinforcement problem and discuss it from a causal perspective. Third, we elaborate on the experimental setup details and present the agent performance results. Finally, we summarize our findings in the conclusion section for this autonomous driving agent study.

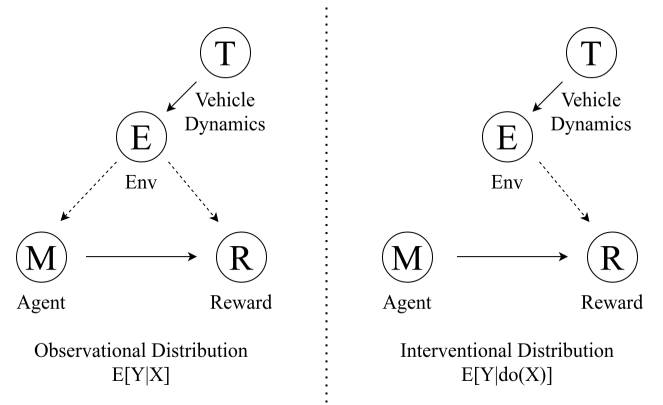


Fig. 2. Observational (*Left*) and Interventional (*Right*) distributions represented with causal graphs.

II. PREVIOUS WORK

In the past Deep-RL have made exhaustive contributions in traffic control tasks [19], [20]. CARLA [21] simulation environments provide a system framework that estimates various affordances from sensors [22]. Solving exploration based driving tasks like vehicle merging in traffic etc. have shown promising results in simulators [23]. ChauffeurNet [24] discusses the advantages of using imitation learning for training robust autonomous agents that leverage worst case scenarios from realistic collision settings for learning.

The highway-env python package focuses on designing safe working policies for nonlinear autonomous driving systems [14]. This package has been extensively used for modelling different types of MDPs, for example: finite stochastic MDP, constraint-MDP and budgeted-MDP (BMDP) [25]. The BMDP variant ensures that the risk notion implemented as cost signal stays below a custom threshold value [26]. Already problem formalization mechanisms exist for defining workings of temporal abstraction, partial observability, reward hypothesis and vehicle kinematics. The BMDP safety and efficacy study extends to continuous kinematics states and unknown human behavior from finite state space and known dynamics [26]. Model free learning techniques that approximate these MDPs like Deep Q-Learning & Dueling Deep Q-Learning networks have shown promising results in continuous agent learning [27], [28].

Collision scenario knowledge has been leveraged for model based algorithms by high confidence region building consisting of true dynamics with high probability. Robust stabilisation and mini-max control with generic costs was achieved by using Tree based planning algorithms. Also, the same studies used interval prediction and non-asymptotic linear regression for safer trajectory predictions [14], [25]. Behavior guided action research which uses safety trajectory computes and proximity graphs to work with aggressive and conservative scenarios have also shown promising results [29]. The CMetric study was used to generate varying degrees of aggressiveness in traffic for autonomous agent behavior analysis [30]. But, in

our case we use more randomization and traffic clogging at key critical spaces across multiple lanes to evaluate agents results at a more granular level.

Reason based interpretations have been extensively provided by causal modelling techniques for many different domains and use-cases [31]–[33]. Additionally, randomized control trials (RCTs) by Fischer are the gold standard for causal link discovery from observational data in any research design experiment [17]. And Sewall’s path diagrams provide a mathematical framework for causal answers [34]. Currently, causal diagrams and multiple adjustments on these diagrams proffer causal relation information for any experimental variable under study [16], [17], [35], [36]. In our experimental setup we utilize these tools to infer direct causal robustness improvements from our environment interventions during agent training.

III. PROBLEM FORMULATION

We intend to measure the performance improvement impact causally with the learnings from collision prone situations for different traffic scenarios. A MDP outputs action a for given state s for any given behavior policy $\pi(a|s)$. From the action set of (*left*, *right*, *break*, *accelerate*, *idle*) selection values for next time step for driving tasks our goal is to learn this behavior policy to predict safe and efficient trajectories. The state-action value function $Q^\pi(s, a)$ for given (s, a) pair estimates future rewards of given behavior policy π . Hence, the optimal state-action value function $Q^*(s, a)$ provides maximum value function estimates for all $s \in \mathcal{S}$ and is evaluated by solving the Bellman Equation [39], provided below for reference.

$$Q^*(s, a) = \mathbb{E}[R(s, a) + \gamma \sum_{s'} P(s'|s, a) \max_{a'} Q^*(s', a')] \quad (1)$$

Traffic flow in the highway-env gets generated from MOBIL model [38] for lane changing & Intelligent Driver Model (IDM) [37] for linear acceleration. First criterion determines the net advantage of lane changing in defined terms of complete acceleration gain and second safety criterion checks whether post lane change a given vehicle is having acceleration room to drive. Our model free learning architecture consists of dueling DQN networks encompassing value and advantage streams which approximate $Q(s, a)$ function to predict best possible actions from the policy π .

We intend to directly measure the impact of adding collision scenario interventions in driving environments. Our evaluation setup consists of calculation of average reward values against perturbed and control environments randomly which makes this setup equivalent to RCT. In this setup the backdoor criterion is satisfied meaning that there is no unobserved confounding variable present here. We know that in RCTs distribution of all covariates are the same except the treatment. This further implies that association is equal to causation here as covariate balance in observational data is present. Meaning that we can use the associative difference quantity to calculate treatment effect on potential outcomes. Finally, we can use the Average Treatment Effect (ATE) approach to evaluate causal

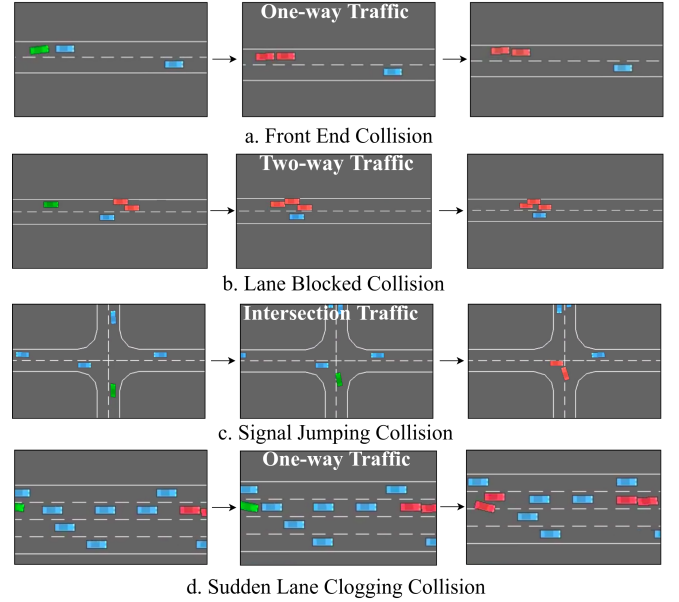


Fig. 3. Probable collision example simulations generated for different driving tasks.

performance improvements by subtracting averaged out values of treatment and control potential outcomes.

Information about the system dynamics is not generally available and model free learning approaches accumulate learning only from environment interactions. Therefore, DQN based autonomous agents generalizes policies only from the particular scenarios encountered during the model training. Figure 3 demonstrates different real-life collision situations possible in different driving tasks. The cases *a* and *c* demonstrate sudden brake collision scenarios where our ego-vehicle collides with suddenly stopped vehicles in traffic. In cases *b* and *d*, collision because of clogged ego-vehicle trajectory by already collided and moving traffic is highlighted respectively. These cases present few of the real life collision scenarios where conservatively trained model-free agents fail. For increasing the critical case coverage of the autonomous agent we add risky randomized behavior during the model training which in turns introduces collision based corner cases for learning. This further increases capacity to handle critical scenarios to avoid hefty collisions. Further, we causally analyze the performance improvements of our treatment agents in comparison to control ones.

Geometrically intersection of two convex polygons rendered in the highway-env package represents the collisions between the vehicles. And these collisions get detected with the separating axis theorem mathematically. The main concept is to find a separating line for the 2D polygons and if that line is found the collision hasn’t happened yet. Essentially, for each given edge of our polygon we find a perpendicular axis to the current edges under review. Post that we project these edges along these perpendicular axes and check whether the projections don’t overlap with the other polygon, refer Figure 4. In essence it functionally means that no collision has happened yet.

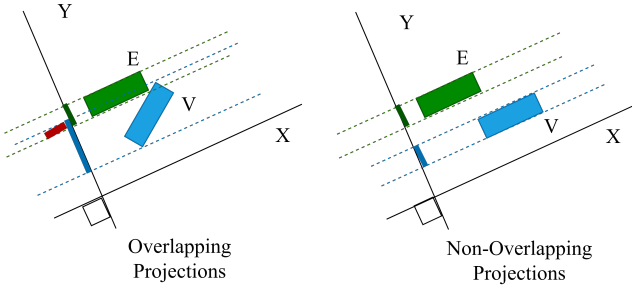


Fig. 4. Overlapping (Left) and Non-Overlapping (Right) projection analysis for collision detection in highway-env package.

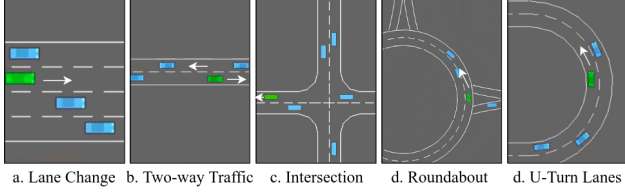


Fig. 5. Different driving navigation environment tasks for collision prone scenario analysis.

IV. EXPERIMENTATION

We evaluate the ATE metric in our experimentation setup for five tasks: roundabout, intersection, lane changing, two-way traffic and u-turn as shown in Figure 5. These five tasks are evaluated at different specific granularities against increasing vehicle count/density to a 200 % increase from respective default values. For all these environments we create treatment environments with dangerous steering and acceleration parameters that are non-compliant with MOBIL model criteria. We create these collision prone scenarios by significantly changing the max-min acceleration and steering parameters highlighted in equations stated below to create more dangerous driving kinematics rules defined in highway-env package. The addition of this risk factor knowledge during the agent learning process allows us to quantify model performance improvements with our experimentation setup causally. And finally we compare it with the control baseline models across these navigation tasks against a spectrum of increasing traffic density.

$$\begin{aligned} acc_p &= acc_{min} + rand[0,1] * (acc_{max} - acc_{min}) \\ str_p &= str_{min} + rand[0,1] * (str_{max} - str_{min}) \end{aligned} \quad (2)$$

We rebuild the highway-env package environments with our custom traffic configurations each time for given specific driving scenarios. Every newly built treatment environment specific to every driving task consists of increased traffic density and more riskier randomized behavior. And their counter control environments only contain increased traffic density but not the randomized collision prone behavior. For example in the lane changing task of episode duration 20 seconds, in control and treatment agent training we incrementally move the vehicle count from 50 to 150 with equivalent intermittent increase of vehicle density by 100 %. We train distinct agents

for navigation on every different traffic count environment version in our experiment corresponding to every vehicle interval count increase. Also, we summarize our ATE findings as a comparative analysis plot of treatment and control agents across these different environment iterations.

In a similar manner, we uniformly increase vehicle count from 3 to 12 with iterative increments of 3 traffic vehicles for the u-turn tasks. These vehicles are added uniformly across the two forward moving lanes on the critical turning area. And while designing for the two-way traffic environment we first reduce the original environment length to 2/3rd. Post that we iteratively increment the vehicle count from 5 in direction & 2 opposite direction to 15 in direction & 6 opposite direction in two more steps altogether. For collision scenario treatment corresponding to the intersection task we add more risk by randomizing the acceleration and deceleration even further more. In addition to that we increase the vehicle from 10 to 30 incrementally with an interval gap of 5 vehicles plus increase the spawning probability by 0.1 iteratively till it reaches maximum threshold. Finally, for the roundabout task we increase the traffic from 5 to 15 vehicles with more risky randomized driving behavior encoded into our treatment environments for agent training and performance comparison with control agents.

All the above traffic configurations require the treatment and control environments to be rebuilt every time for repeated agent training. To evaluate each such agent from both control and treatment environment configurations we kept the traffic as constant in our test population set. In order to be compliant with our causal experiment structure we only changed risky behavior parameters in treatment environment configurations for ATE calculations corresponding to a given traffic population setting.

We use the Dueling DQN model architecture with ADAM optimizer having learning rate 5e-4 in our experiments during agent training. The discount factor is inputted as 0.99 and observational agent interaction data is fed in the batch size of 100. Our agents are trained till the maximum count of 3072 episodes until they converge to average reward value specific to that environment. We utilize the dueling network design which uses advantage function $A(s,a)$ to assist in estimation of $Q(s,a)$ function for state action pairs more accurately. This architecture splits the network into two streams, namely value and advantage ones which contain some base hidden layers in common. As shown in Figure 6 the shared network has 3 fully connected layers of 256, 192 and 128 units respectively and the final output of these streams is fully connected to the remaining network. The value stream calculates the value function for a given state with its one output. And the advantage stream consists of n_a outputs corresponding to all discrete possible actions for a given state. These two outputs are finally combined in the dueling architecture to get the $Q(s,a)$ function values for state-action pairs, given by $Q(s,a) = V(s) + A(s,a)$ formula.

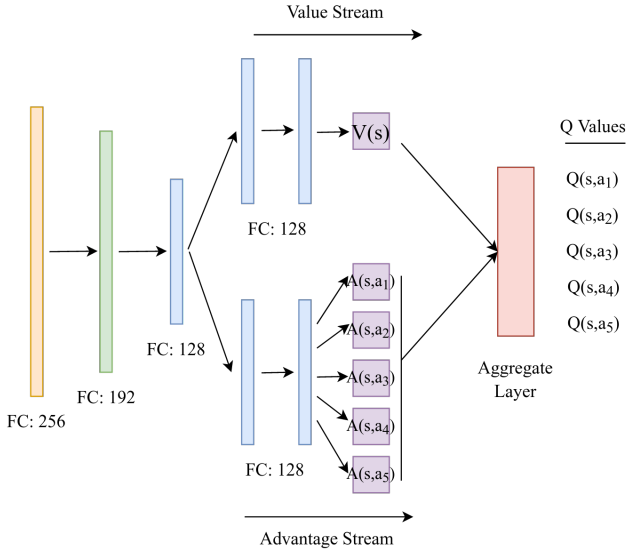


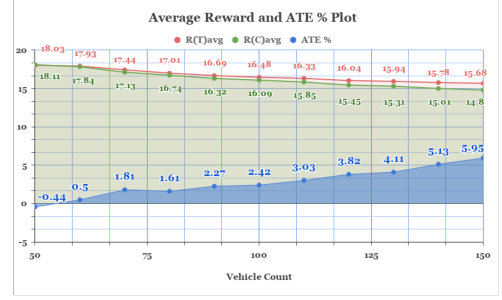
Fig. 6. DQN based dueling model architecture used for autonomous agent training.

V. RESULTS

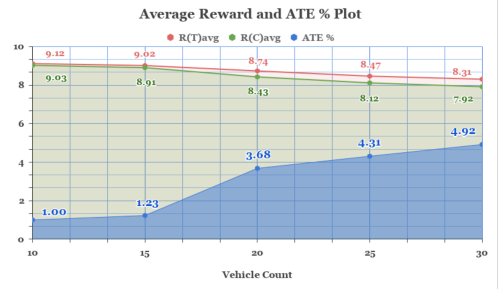
For uniformly adding randomization across different environments corresponding to the same driving scenario we select different initial random seed values. This essentially creates variations in environment components like vehicle location, spawning vehicles count, vehicle acceleration & deceleration etc. With different random seed values we average out any anomalous results for comparing ATE of our treatment and control models which is evaluated against collision prone and regular environments in the test set. The ATE calculation used for this experimentation is expressed by the equation 3 where the first two terms calculate the average reward gains achieved by the treatment models in risk prone and regular environments. And the last two terms calculate the same for control agents in both these types of environment samples as well, where $N_T = N_C = 100$ in the final equation.

$$ATE = 1/N_{T,C} * \left[\sum_{i=1}^{N_T} R_{avg^i}^T + \sum_{i=1}^{N_C} R_{avg^i}^T - \sum_{i=1}^{N_T} R_{avg^i}^C + \sum_{i=1}^{N_C} R_{avg^i}^C \right] \quad (3)$$

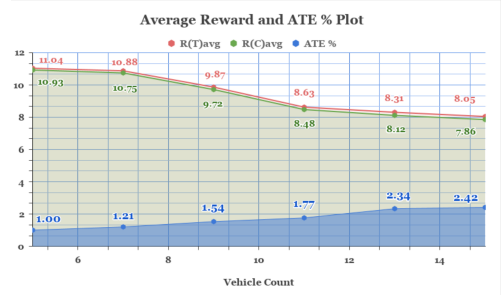
Percentage improvements of ATE results in Figure 7 demonstrates the advantage of learning from collision prone scenarios. These improvements are even more pronounced when the traffic density increases for driving environments like in any real-life situation. There is also a declining trend of average reward values as traffic increases over all the driving scenarios. This depreciation can be attributed to slow progression of the agent across the driving environments. The increase in relative improvements of ATE values as the traffic continues to grow showcases the robustness of treatment models over the control ones. Hence, our causal effect estimation experiment



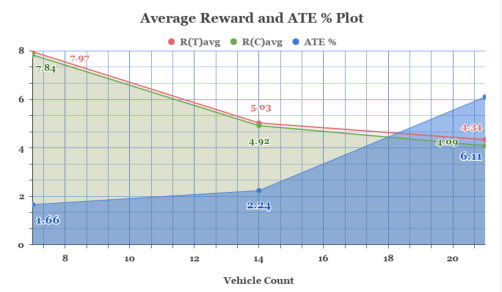
a. Lane Change Task



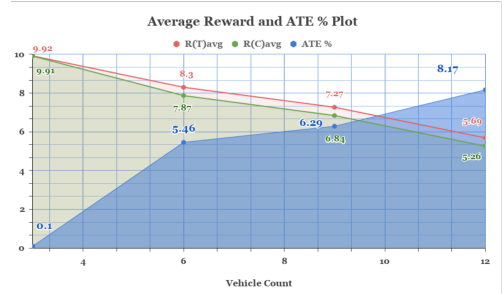
b. Intersection Task



c. Roundabout Task



d. Two-way Task



e. U-Turn Task

Fig. 7. Percentage performance improvement ATE plots for above stated different driving scenarios.

setup quantifies the information learnt from the doped collision prone scenarios. This approach holds good for creating key metrics oriented systems for developing autonomous driving agents.

VI. CONCLUSION

With our causal experimentation setup we generated more rugged agents on the discussed critical driving situations by using the same MDP formulation as earlier. This also proved our initial experimentation hypothesis that agents trained on systematically added critical driving scenarios excels in performance as compared to agents learning from safer driving simulations. For all the discussed environments our agents proved to be better collision deterrents demonstrating the importance of learning from corner cases generated by collision prone scenarios. Here, with our analysis we successfully demonstrated a methodology that systematically measures the agent's ability to learn from risky driving situations.

REFERENCES

- [1] D. Elliott, W. Keen, and L. Miao, "Recent advances in connected and automated vehicles," *Journal of traffic and transportation engineering (English edition)*, vol. 6, no. 2, pp. 109–131, 2019.
- [2] S. A. Bagloee, M. Tavana, M. Asadi, and T. Oliver, "Autonomous vehicles: challenges, opportunities, and future implications for transportation policies," *Journal of modern transportation*, vol. 24, no. 4, pp. 284–303, 2016.
- [3] J. Joy and M. Gerla, "Internet of vehicles and autonomous connected car-privacy and security issues," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–9.
- [4] J. M. Anderson, K. Nidhi, K. D. Stanley, P. Sorensen, C. Samaras, and O. A. Oluwatola, *Autonomous vehicle technology: A guide for policymakers*. Rand Corporation, 2014.
- [5] L. Li, W.-L. Huang, Y. Liu, N.-N. Zheng, and F.-Y. Wang, "Intelligence testing for autonomous vehicles: A new approach," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 2, pp. 158–166, 2016.
- [6] M. Koren, S. Alsaif, R. Lee, and M. J. Kochenderfer, "Adaptive stress testing for autonomous vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1–7.
- [7] L. Cui, J. Hu, B. B. Park, and P. Bujanovic, "Development of a simulation platform for safety impact analysis considering vehicle dynamics, sensor errors, and communication latencies: Assessing cooperative adaptive cruise control under cyber attack," *Transportation research part C: emerging technologies*, vol. 97, pp. 1–22, 2018.
- [8] P. Liu, Z. Xu, and X. Zhao, "Road tests of self-driving vehicles: affective and cognitive pathways in acceptance formation," *Transportation research part A: policy and practice*, vol. 124, pp. 354–369, 2019.
- [9] L. Li, X. Wang, K. Wang, Y. Lin, J. Xin, L. Chen, L. Xu, B. Tian, Y. Ai, J. Wang et al., "Parallel testing of vehicle intelligence via virtual-real interaction," *Science robotics*, vol. 4, no. 28, p. eaaw4106, 2019.
- [10] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," *Nature communications*, vol. 12, no. 1, pp. 1–14, 2021.
- [11] W. Ding, B. Chen, M. Xu, and D. Zhao, "Learning to collide: An adaptive safety-critical scenarios generating method," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2243–2250.
- [12] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning visual classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1625–1634.
- [13] C. Xie, J. Wang, Z. Zhang, Y. Zhou, L. Xie, and A. Yuille, "Adversarial examples for semantic segmentation and object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1369–1378.
- [14] E. Leurent, Y. Blanco, D. Efimov, and O. A. Maillard, "Approximate robust control of uncertain dynamical systems," *arXiv preprint arXiv:1903.00220*, 2019.
- [15] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.
- [16] J. Pearl et al., "Causal inference in statistics: An overview," *Statistics surveys*, vol. 3, pp. 96–146, 2009.
- [17] J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect*. Basic books, 2018.
- [18] P. W. Holland, "Statistics and causal inference," *Journal of the American statistical Association*, vol. 81, no. 396, pp. 945–960, 1986.
- [19] F. Belletti, D. Haziza, G. Gomes, and A. M. Bayen, "Expert level control of ramp metering based on multi-task deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1198–1207, 2017.
- [20] C. Wu, A. Kreidieh, E. Vinitsky, and A. M. Bayen, "Emergent behaviors in mixed-autonomy traffic," in *Conference on Robot Learning*. PMLR, 2017, pp. 398–407.
- [21] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
- [22] A. Sauer, N. Savinov, and A. Geiger, "Conditional affordance learning for driving in urban environments," in *Conference on Robot Learning*. PMLR, 2018, pp. 237–252.
- [23] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "Safe, multi-agent, reinforcement learning for autonomous driving," *arXiv preprint arXiv:1610.03295*, 2016.
- [24] M. Bansal, A. Krizhevsky, and A. Ogale, "Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst," *arXiv preprint arXiv:1812.03079*, 2018.
- [25] E. Leurent, "Safe and efficient reinforcement learning for behavioural planning in autonomous driving," Ph.D. dissertation, Université de Lille, 2020.
- [26] N. Carrara, E. Leurent, R. Laroche, T. Urvoy, O.-A. Maillard, and O. Pietquin, "Budgeted reinforcement learning in continuous state space," *arXiv preprint arXiv:1903.01004*, 2019.
- [27] S. Gu, T. Lillicrap, I. Sutskever, and S. Levine, "Continuous deep q-learning with model-based acceleration," in *International Conference on Machine Learning*. PMLR, 2016, pp. 2829–2838.
- [28] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 1995–2003.
- [29] A. Mavrogiannis, R. Chandra, and D. Manocha, "B-gap: Behavior-guided action prediction for autonomous navigation," *arXiv preprint arXiv:2011.03748*, 2020.
- [30] R. Chandra, U. Bhattacharya, T. Mittal, A. Bera, and D. Manocha, "Cmetric: A driving behavior measure using centrality functions," *arXiv preprint arXiv:2003.04424*, 2020.
- [31] J. Hicks et al., *Causality in economics*. Australian National University Press, 1980.
- [32] D. A. Smirnov and I. I. Mokhov, "From granger causality to long-term causality: Application to climatic data," *Physical Review E*, vol. 80, no. 1, p. 016208, 2009.
- [33] C. W. Granger, "Some recent development in a concept of causality," *Journal of econometrics*, vol. 39, no. 1-2, pp. 199–211, 1988.
- [34] S. Wright, "The method of path coefficients," *The annals of mathematical statistics*, vol. 5, no. 3, pp. 161–215, 1934.
- [35] T. S. Richardson and J. M. Robins, "Single world intervention graphs: a primer," in *Second UAI workshop on causal structure learning*, Bellevue, Washington. Citeseer, 2013.
- [36] E. Bareinboim, J. Correa, D. Ibeling, and T. Icard, "On pearl's hierarchy and the foundations of causal inference," *ACM Special Volume in Honor of Judea Pearl (provisional title)*, 2020.
- [37] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.
- [38] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model mobil for car-following models," *Transportation Research Record*, vol. 1999, no. 1, pp. 86–94, 2007.
- [39] R. E. Bellman and S. E. Dreyfus, *Applied dynamic programming*. Princeton university press, 2015, vol. 2050.